

The influence of reputation information on the assessment of undergraduate student work

Assessment & Evaluation in Higher Education

John Batten^{a*}, Jo Batey^a, Laura Shafe^a, Laura Gubby^b and Phil Birch^c

^aDepartment of Sports Studies, The University of Winchester, Winchester, UK; ^bSport Science, Tourism and Leisure, Canterbury Christ Church University, Kent, UK; ^cSports, Exercise and Health Sciences, University of Chichester, Chichester, UK

The present study employed an experimental design to examine the influence of knowledge of a student's previous performance and the general quality of their writing style on the assessment of undergraduate student work. Fifteen sport and exercise physiology academics were asked to mark and give feedback on two final year undergraduate student essays. The first student essay that participants marked was a control essay. The second essay was the target essay. Participants read one of three student reputation profiles (positive, negative or neutral) prior to marking this essay. Kruskal–Wallis tests for difference indicated that the marks awarded to each essay did not significantly differ between the three student reputation conditions. Thematic analysis of the target essay also revealed no apparent differences in the way in which feedback was presented across the three student reputation profiles. It was therefore concluded that non-anonymous marking did not induce marker bias in this instance.

Keywords: assessment; anonymity; bias; pedagogy

Introduction

The implementation of anonymous marking, whereby student identity is withheld from the assessor as a means to eliminate bias, is a persistent and controversial concern within higher education (Whitelegg 2002; Brennan 2008; Owen, Stefaniak, and Corrigan 2010). Indeed, this contested issue not only divides individual academics, but there is also disparity in the practice of anonymous marking across the entire higher education sector. Research that has endeavoured to clarify the effectiveness of anonymous marking has also produced equivocal results; with some finding that anonymous marking could eliminate bias (e.g. Bradley 1984), whilst others failed to identify any real need for anonymous marking (e.g. Newstead and Dennis 1990).

The perceived discrimination which materialises from knowledge of the student's demographic status has also been the focus of previous investigations (e.g. Spear 1984; Bradley 1993; Newstead and Dennis 1993; Dennis and Newstead 1994; Baird 1998), whereas the potential for personal knowledge of the student's previous performance to provoke bias in marking is yet to be thoroughly explored (Huot 1990). Only when educational institutions are more aware of the potential implications of such additional sources of bias will they be able to make more

informed and justified decisions about their marking practices, and thus, defend the integrity of student assessment. Consequently, the aim of this study was to examine the influence of personal knowledge of a student's previous performance and the general quality of their writing style on the assessment of undergraduate student work.

Theoretical underpinning

The interactions that take place during lectures, seminars, workshops and/or tutorials provide the social grounds from which a lecturer can gather and integrate information to form an assumed holistic account of an individual student, otherwise known as an impression. However, a lecturer's experiences of a particular student are often compounded with stereotypes, and together they may help to consolidate, reinforce or alter the impressions a lecturer makes (Jussim 1986). Those at a disadvantage, therefore, are students whose lecturers have formed an erroneous impression of them, or those who are tarnished by their lecturers as poor performers. Yet, impressions alone cannot account for the underlying processes responsible for bias in assessment (Fiske and Taylor 1991). Instead, it is the interaction between impressions and expectations which will determine a lecturer's behaviour.

Fiske and Taylor (1991) contend that the impressions people make about others will pave the way for their expectations. More specifically, it is argued that predictions will be made about the target's (i.e. the student's) future behaviour based on the initial information that the perceiver (i.e. the lecturer) has access to (Brophey 1983; Hilton and Darley 1991). The significance of expectations in the context of education was first realised following research within primary schools. It was concluded that teachers' expectations could potentially influence their behaviour towards the student, to the extent that the teacher would seek to verify their expectation by eliciting confirmatory behaviours in the student (Braun 1976; Dipboye 1985). This phenomenon is known as a self-fulfilling prophecy (Merton 1948), which proposes that '... one person's expectations about a second person leads the second person to act in ways that confirm the first person's original expectation' (Jussim 1986, 429).

However, in the case of student assessment, where there is no direct social interaction between the perceiver and target, it still remains possible that a lecturer may attempt to fulfil the prophecy they have previously laid out for the student. According to Jussim (1986), teacher expectations can influence the way in which a student's performance is interpreted, to the extent that it yields confirmatory information. For instance, when marking an assignment, a lecturer may see qualities in an essay which complement their expectation, but are not in fact present or valid (Huot 1990). Jussim (1989) and Jussim and Eccles (1992) refer to such occurrences as perceptual biases; whereby a student is viewed as performing more consistently with the teacher's expectation than is actually deserved, reflected in the grades awarded. Although this research was based on classroom settings in primary education, the same principles may be applied to marking practices within the higher education context.

The theory of cognitive dissonance (Festinger 1957) provides further insight into the matter of perceptual biases. It proposes that when individuals are exposed to information that contrasts with their beliefs and expectations, an unpleasant psychological state is experienced which they seek to resolve. As a consequence,

the lecturer is more inclined to simply discount or devalue the opposing evidence, as a means to reduce the dissonance, rather than to re-evaluate the impressions and expectations already made (Braun 1976; Brennan 2008). Thus, the lecturer places greater importance and pays more attention to consistent aspects of, for example, a student's essay, than to the elements that oppose what they have come to expect from the student.

For instance, when a student is expected to produce poor-quality work but outperforms this initial expectation on one assignment, they may not be given adequate credit because the teacher's expectation does not allow for it. Alternatively, students for whom lecturers hold high expectations are likely to be given the benefit of the doubt when they underachieve (Jussim 1986; Ecclestone 2001). This is not to say that all expectations will lead to biased practices, as it will depend on the strength and flexibility of the expectation, along with the details of the opposing evidence, such as its frequency relative to expectancy-consistent information (Jussim 1986). The nature of the assessment will also mediate the extent to which the teacher's expectation can influence the judgements made; whereby the more subjective the assessment and criteria, the more room there is for biases to operate (Archer and McCarthy 1988; Dennis, Newstead, and Wright 1996).

However, and notwithstanding the comments made above, some lecturers may still be unwilling to modify their impressions and/or expectancies due to the high cognitive demand that this would likely require. More specifically, the continuum model (Fiske and Neuberg 1990) contends that an individual may use either a schema-driven or data-driven information-processing approach, depending on whether or not they have sufficient cognitive resources to attend to the information at hand. Under conditions of high cognitive load, Snyder and Stukas (1999) posit that perceivers will attempt to manage the task of interpreting information by placing an increased reliance on their expectancies; as opposed to attending to individual information. Consistent with this notion, Plessner (2005) found that when a decision needs to be reached quickly, and when the time demands of the situation restrict the processing of all available information, schema-driven information processing is more likely to be used. Bargh and Thein (1985) also found that the ability to engage in a more data-driven information-processing approach is dependent on the availability of sufficient cognitive capacity.

With particular reference to marking student papers, schema-driven theorists (e.g. Fiske and Taylor 1991) would argue that a lecturer assigns a student to a specific category, for example, good student or bad student, based on those cues available either before an interaction or in the early stages of an interaction. These schemas then enable a lecturer to make a judgement about the characteristics and mental states of a student, for instance, good students are industrious, bad students are lazy, and to form expectancies for the interaction. Schemas also have the potential to influence a lecturer's information processing and their affective responses to a student. This is done by influencing what information is attended to, how that information is encoded and evaluated, and the information that is remembered (Chapman and Chapman 1967; Higgins and Bargh 1987). Therefore, schemas may impact the marking process by leading a lecturer to think and act in such a way as to cause their initial expectancy to come true. This process is typified by the self-fulfilling prophecy phenomenon.

Data-driven theorists (e.g. Anderson 1981), on the other hand, would question the extent to which a lecturer's initial expectancies would influence the marks they

award. More specifically, data-driven theorists would argue that a lecturer forms an impression by integrating every new piece of information in a systematic and unbiased fashion. If this is true, then their initial expectancies will have a limited impact on the marking process. Olson, Roese, and Zanna (1996) also state that disconfirmation of an expectancy (e.g. when a student who is expected to produce poor-quality work outperforms this expectation) will instigate a greater systematic analysis of the presented information. Indeed, the surprise experienced when a student behaves in a way that is inconsistent with a lecturer's initial expectancy should make their original expectancy more salient, which in turn should encourage them to pay more attention to their initial prediction. In contrast to the propositions of Braun (1976), this might be an alternative way for a lecturer to resolve the dissonance which emanates when they are exposed to information that contrasts with their beliefs and expectations.

Previous research

A wide variety of sources, including the student's ethnicity, socio-economic background, and age, as well as physical attractiveness, can all potentially induce erroneous judgements in the minds of lecturers (Braun 1976; Archer and McCarthy 1988; Meadows and Billington 2005). However, the majority of research to date has explored the impact of the student's gender on the assessment process. In particular, early investigations attempted to dissect the pattern that had emerged in grade distribution, whereby male students tended to receive more extreme degree classifications (1st or 3rd class) and females were often awarded 2nd class degrees (Newstead 1996; Francis, Robson, and Read 2001). Thus far, however, findings have been equivocal in determining the extent and manner in which gender bias operates.

For example, Bradley (1984) investigated the differences in the marks awarded to final-year projects between a student's supervisor and a second marker, who presumably had less personal knowledge of the student. It was hypothesised that gender bias would occur in the second marker, whereas the supervisor would be more in touch with the student's true ability. Not only was this hypothesis accepted, but additional data confirmed that when the projects were anonymously assessed by the second marker, differences between the two markers were no longer significant. Bradley (1984) therefore concluded that blind marking eliminated gender bias. Newstead and Dennis (1990), on the other hand, found no significant differences in the grades awarded by a supervisor and second marker, leading to the conclusion that gender bias was not present in this instance.

Dennis, Newstead, and Wright (1996) later used structural equation modelling to analyse the marks awarded to student projects. They found that approximately 30% of the variance in the marks arose from factors that influenced the supervisor but not the second marker; with the most likely factor being the supervisor's personal knowledge of the student. This is consistent with the research by Hand and Clews (2000), who conducted a focus group interview with undergraduate tutors, finding that many believed supervisors to give higher marks to dissertations than did second markers, partly due to their experience of the tutee during the supervision process. In line with this contention, Ecclestone (2001) observed staff to 'compensate' for the assessment criteria in order to take into account their perception of the student's application, conscientiousness, personal pressures, personal progress and contributions during tutorials.

However, experimental research that has manipulated the lecturers' knowledge of a student (i.e. reputation) to examine its impact on the marks awarded to student projects is limited. For example, Dennis, Wright, and Rigsby (1987) conducted an experiment in which students were assigned to different groups based on their reputation. They found that lecturers awarded higher marks to students with a good reputation than to those with a poor reputation. This is an Accepted Manuscript of an article published by Taylor & Francis in *ASSESSMENT & EVALUATION IN HIGHER EDUCATION* on 9 December 2014, available online at: <https://www.tandfonline.com/doi/abs/10.1080/02602938.2011.640928>.

found that the same essays were marked higher when they were believed to have been produced by competent students. Fleming (1999) also found that tutors tended to award higher marks to students with previous track records of good grades. It is therefore logical to consider that a student's previous performance may be perceived by a lecturer as an accurate predictor of future assessment outcomes. As a consequence, students with a good reputation based upon excellent performance in previous assessments might be at an advantage when anonymous marking is not implemented, whereas those who are known for poor performance may find their reputation prevents them from achieving the high grades they actually deserve.

The positive consequences of lecturer expectations are known as 'Galatea Effects', whereas the negative consequences are known as 'Golem Effects' (Babad, Inbar, and Rosenthal 1982). Ethically, studying the positive consequences of favourable lecturer expectations (i.e. the 'Galatea of the classroom') is more acceptable than studying the negative consequences of lecturer expectations. However, in terms of ecological validity, it is equally important to study the 'Golem of the classroom' as well (Babad, Inbar, and Rosenthal 1982). Moreover, research examining the consequences of lecturer expectancies has often reported equivocal results in relation to the strength of both positive and negative expectations. Indeed, whilst Babad, Inbar, and Rosenthal (1982) found no consistent trends regarding the strength of positive and negative expectancies, Sutherland and Goldschmid (1974) identified stronger negative expectancy effects. Conversely, and with the adoption of a larger sample, Madon, Jussim, and Eccles (1997) reported that positive expectancy effects were generally more powerful than negative ones. Thus, the present study sought to examine the consequences of both positive and negative expectancy effects in relation to marking student work in higher education.

The majority of the research reviewed thus far has also maintained a focus on the marks awarded to student work, yet the biases which may exist within the comments provided throughout the written assignment and at the end of each essay have largely been neglected. These aspects of the marking procedure serve to highlight the strengths, weaknesses and necessary future actions that will progress the student towards greater academic achievement (Rust 2002). Previous research by Sadler (2010) has also highlighted the importance of clear assessor feedback as a means of assisting students in their learning and development. As such, future investigation concerning the impact of marker bias on the construction of feedback is clearly warranted. Jussim (1986) adds further weight to this claim by suggesting that the teacher provides clearer feedback to the low-performing high-expectancy students as a means to bring them closer to their expectation, whereas the high-performing low-expectancy students receive less intricate support.

Project aims

At present, there is a lack of empirical evidence concerning the impact of a student's reputation on the marks which they are awarded and the way in which feedback is constructed throughout an essay. As a result, the specific aim of this study was to examine the influence of reputation information, in the form of knowledge of a student's previous performance and the general quality of their writing

style, on the assessment of undergraduate student work. The decision to manipulate information about a student's previous performance and the general quality of their writing style concurrently in this study was based on the findings of previous research. More specifically, Pain and Mowl (1996) and Elander (2002), amongst others (e.g. Elander et al. 2006), have reported these two variables to be inextricably linked. It was therefore hypothesised that those students with a more positive reputation would receive significantly more favourable marks, and would receive more feedback than those students with a negative reputation.

Method

Participants

A total of 15 sport and exercise physiology academics (n males = 10, n females = 5; mean age 38.0, s = 9.93 years) were recruited from eight higher education institutions across England, Scotland and Wales. The participant sample (mean experience in higher education of more than 10 years) represented a total of five academic positions (n heads of department = 1; n readers/principal lecturers = 2; n senior lecturers = 6; n lecturers = 4; n teaching assistants = 2) and reported various marking loads (n less than 50 essays = 4; n 50–100 essays = 3; n 100–200 essays = 3; n 200–500 essays = 4; n 500–1000 essays = 1) across the 2008/2009 academic year. The experimental protocol was explained to the participants and ethical approval and written informed consent obtained.

Materials

Student work

The aforementioned sport and exercise physiology academics were asked to mark and give feedback on the same two final year undergraduate student essays. The sample essays were approximately 2500 words in length and had previously been submitted for assessment in the Sports Studies Department at The University of Winchester. The original assessors (n = 2) confirmed both essays to be of a relatively equal standard (lower second). Consistent with the recommendations of Francis and colleagues (e.g. Francis, Robson, and Read 2001, 2002), second class (2:2) essays (which have been found to contain less gender stylistic markers) were used to minimise the potential for gender bias within the assessment process. Read, Francis, and Robson (2005) also contend that lower second-class essays should stimulate more detailed reflections from the prospective markers as they have both strengths and weaknesses. Written informed consent was obtained from the students to use their original sport and exercise physiology essays for this research.

Assessment criteria profiles

Francis et al. (2003) and Read, Francis, and Robson (2005) found that academics have a tendency to use different criteria (typically from their own institutions) to help them to assess the quality of student work. With this in mind, participants in the present study were asked to use the marking scheme (Assessment Criteria Profile [ACP]) from the Sports Studies Department at The University of Winchester to help to standardise the experimental protocol. The ACP was derived from academic discourse with external examiners and provides students with immediate, descriptive

and diagnostic feedback about key aspects of the assessment; e.g. accuracy and interpretation of work studied, quality and suitability of examples used, maturity and critical thinking, etc.

Procedure

Participants were asked to carefully read a written description (reputation profile) of the student prior to marking each essay, and were equally divided ($n = 5$) into the three reputation conditions (positive, negative or neutral). The reputation profile of each student was portrayed to the participants in such a way so as to emphasise the importance of this information in helping to contextualise the assignment. All of the reputation profiles were adapted from the descriptive information presented by Martin Ginis and colleagues (e.g. Martin, Sinden, and Fleming 2000; Martin Ginis, Latimer, and Jung 2003) and Greenlees et al. (2007).

The first student essay that all participants marked was a control essay. The reputation profile for the control student essay was the same for all of the groups, and was as stated:

Ben is a 22-year-old final year undergraduate Sports Studies student. His work has been of varying standard and he averaged a 2:2 in the first Semester. Ben is enthusiastic about sport and works as a fitness instructor at a local gymnasium. The following essay was submitted for assessment on the Sport and Exercise in Extreme Environments module.

Participants then marked the target essay. All participants marked the same essay, but were required to read one of three student reputation profiles prior to marking this essay. The reputation profiles manipulated within the present study informed the reader of the student's previous performance (i.e. 3rd class, 2:1 or neutral) and the general quality of their writing style. The reputation profile for the 2:1 (positive) student essay was as stated:

Helen is a 21-year-old final year undergraduate Sports Science student. Her writing style is generally very good and she averaged a 2:1 in the first Semester. Helen is passionate about sport and has played competitive netball for 10 years. The following essay was submitted for assessment on the Sport and Exercise in Extreme Environments module.

The reputation profile for the 3rd class (negative) student essay was the same as above, except that the second sentence was altered to read: 'Her writing style is generally very poor and she averaged a 3rd class in the first Semester'. The reputation profile for the neutral student essay was also the same as above, except that the second sentence was omitted; meaning that participants in this condition received no information regarding the student's previous performance or the general quality of their writing style.

Participants were required to mark and give feedback on the essays as if the work was to be returned to the students. However, participants were also informed that they should utilise the ACP to help them to assess the quality of the work. As a result, this study entailed the exploration of: (a) the spread of marks awarded to each essay; and (b) the ways in which feedback was presented both in-text and at the end of each essay.

Data analysis

The assumptions that underpin tests for difference were examined prior to the further exploration of the data-set. However, although the Shapiro-Wilk and Levene's Test results largely satisfied the requirements of normal distribution and homogeneity of variance (Tabachnick and Fidell 2007), the relatively small sample size ($n = 15$) and number of experimental conditions ($n = 3$) employed within this study necessitated the need for non-parametric tests for difference. As a result, two χ^2 Kruskal–Wallis tests for difference were used to examine the impact of reputation information on the overall marks awarded for: (a) the control essay and (b) the target essay. The independent variable was reputation group, and the dependent variable was the overall mark awarded to each essay. In line with the recommendations of Field (2005), the exact significance values of the Kruskal–Wallis tests for difference were examined. Statistical significance was set at $p < 0.05$ and all analyses were computed using the Statistical Package for Social Sciences (SPSS v.16).

Written feedback

In an attempt to facilitate comparisons between the three reputation conditions, a thematic analysis of both the in-text and end-of-text comments for each of the independent conditions was conducted. A second layer of thematic analysis, which involved all of the experimental essays collectively, was then undertaken. In line with the recommendations of Maykut and Morehouse (1994) and Gratton and Jones (2004), a process of peer de-briefing was also engaged in. This involved another researcher scrutinising the audit trail and raising questions of bias where necessary. The aim of this procedure was to ensure that the notions of trustworthiness, endemic in qualitative research, were adhered to. The two researchers who undertook the thematic analysis also had 15 years of experience in both conducting and publishing qualitative research.

Results

Grade variability for the control essay

The Kruskal–Wallis test for difference revealed no significant difference ($H_{(2)} = 0.564$, $p = 0.782$, $p > 0.05$) in the overall marks awarded to the control student essay across reputation conditions (see Figure 1).

The results for grade variability for the control student essay indicate that the perceptions of the participants did not significantly differ when marking the same essay. This finding enhances the probability that any differences in the perceptions of the target essay are due to the manipulation in reputation information, as opposed to individual differences.

Grade variability for the target essay

The Kruskal–Wallis test for difference revealed no significant difference ($H_{(2)} = 2.545$, $p = 0.291$, $p > 0.05$) in the overall marks awarded to the target student essay across reputation conditions (see Figure 2).

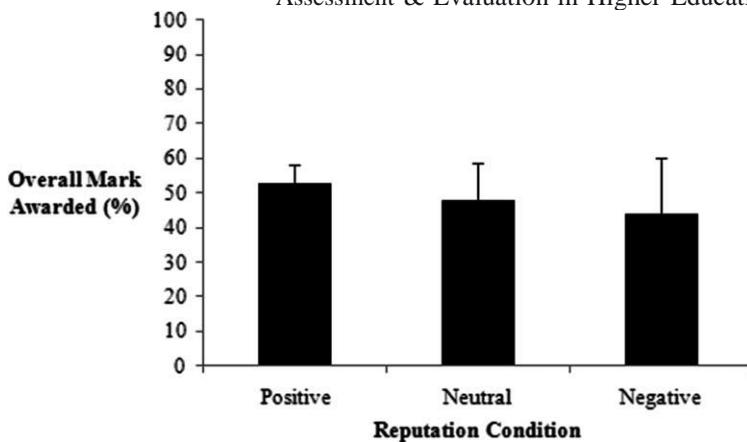


Figure 1. Mean (s) marks awarded to the control student essay across reputation conditions.

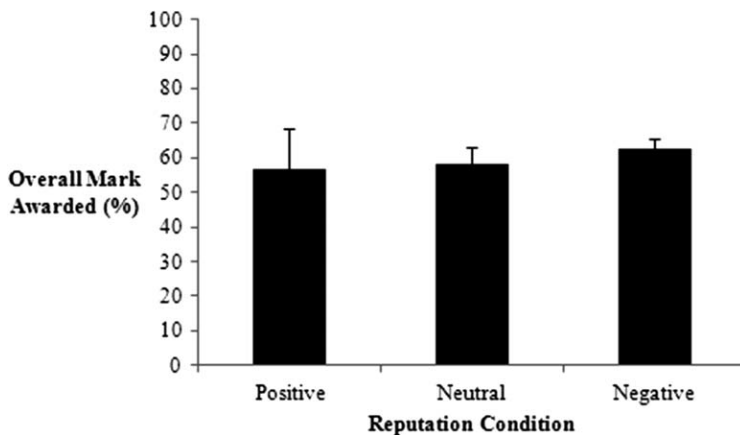


Figure 2. Mean (s) marks awarded to the target student essay across reputation conditions.

Written feedback for the target essay

The themes that emerged from the analysis of the target student essay, and some exemplar data which fall under such themes, are shown in Table 1.

Having conducted a thematic analysis of the target essay, and then compared the identified themes across the positive, neutral and negative reputation conditions, there would appear to be very little difference in how the feedback was presented throughout each essay. More specifically, the criteria on which the markers commented were generic and did not differ between the three reputation conditions. The way in which the feedback was presented, in terms of how animated or emotive it appeared, was also not specific to the type of reputation profile with which the markers were presented. However, the total number of comments that were made on the negative student essay would appear to be higher than on the other two student reputation profiles. In addition, there would appear to be an imbalance between

Table 1. Qualitative analysis of the target student essay across reputation conditions.

Theme	Reputation profile	Total number of comments		Example comments	
		Strengths of the assignment	Areas of improvement	Strengths of the assignment	Areas of improvement
Academic style of writing	Positive	1	1	Well written	Descriptive
	Neutral	0	1	No comments made	Descriptive
	Negative	4	5	Very well written and detailed essay	Avoid commentary style of writing
Criticality	Positive	1	2	Attempted to critically analyse	Attempted to evaluate – but more needed
	Neutral	1	3	Evidence of a critical approach	Lack of comparison and evaluation of studies
	Negative	1	1	The comparison of adults to children is excellent	You could make more comparisons between studies..
Structure, fluency and cohesion	Positive	2	1	Well organised	... overly complex and ... the work loses flow
	Neutral	0	4	No comments made	Lacks flow
	Negative	4	1	Well organised on the whole	Writing style needs to develop more of a flow ...
Sources used	Positive	0	1	No comments made	Dated literature
	Neutral	0	3	No comments made	Numerous secondary references
	Negative	1	2	Good reference list	Where possible try to use primary literature ...
Understanding/knowledge of the subject	Positive	1	1	Well researched	More examples needed
	Neutral	1	1	Well researched	Factual inaccuracies
	Negative	4	3	Good information on mechanisms of heat loss/gain	Some concepts needed further explanation
Other	Positive	0	0	No comments made	No comments made
	Neutral	0	1	No comments made	Lack of planning
	Negative	2	0	Overall, well tried	No comments made

the number of comments made about the relative strengths and areas of improvement within each essay and across the three reputation conditions.

Nonetheless, the following quotations from the 'structure/fluency/cohesion' theme provide a clear example of how the feedback was generally not related to the profile of the student. For instance, a participant who marked the positive student essay commented that 'Too many sentences which are overly complex and consequently the work loses flow', and a participant who marked the negative student essay commented that '... current style is somewhat "broken" and reads as a list in places'. A participant who marked the neutral student essay also commented that 'In this essay you seemed to wander off in various directions without making clear, as you could, why that happened'. When taken collectively, these quotations would seem to indicate that the reputation profile of the student did not influence the way in which the feedback was presented.

Discussion

The specific aim of this study was to examine the influence of reputation information, in the form of knowledge of a student's previous performance and the general quality of their writing style, on the assessment of undergraduate student work. It was hypothesised that those students with a more positive reputation would receive significantly more favourable marks, and would receive more feedback than those students with a negative reputation. The results of the present study, however, are not only in contrast to the proposed hypotheses, but would also appear somewhat contradictory to the results of previous research.

For example, both Diederich (1974) and Rigsby (1987) found that the same essays were marked higher when they were believed to have been produced by competent students. Fleming (1999) also observed tutors to award higher marks to students with previous track records of good grades. However, in contrast to the proposed hypothesis, the results of the present study failed to find any significant differences in the overall marks awarded to students across the three reputation conditions. The total number of feedback comments that were made on the negative student essay were also higher than on the other two student reputation profiles. The remainder of this paper will therefore examine the possible reasons behind these conflicting results and the potential implications of such findings.

Analysis of theory and research

When it comes to marking student papers, proponents of schema-driven theory (e.g. Fiske and Taylor 1991) would argue that a lecturer assigns a student to a specific category based on those cues available either before an interaction or in the early stages of an interaction. These schemas then enable a lecturer to make a judgement about the characteristics and mental states of a student and to form expectancies for the interaction. Schemas also have the potential to influence what information is attended to, how that information is encoded and evaluated, and the information that is remembered (Chapman and Chapman 1967; Higgins and Bargh 1987). Schemas may therefore impact the marking process by leading a lecturer to think and act in such a way as to cause their initial expectancy to come true. This process is typified by the self-fulfilling prophecy phenomenon.

Data-driven theorists (e.g. Anderson 1981), on the other hand, would argue that a lecturer forms an impression by integrating every new piece of information in a
 This is an Accepted Manuscript of an article published by Taylor & Francis in ASSESSMENT & EVALUATION IN HIGHER EDUCATION on 05 December 2011, available online: <https://www.tandfonline.com/doi/abs/10.1080/02602938.2011.640928>.

Although these findings were observed within judging in gymnastics, the typical university lecturer also has a limited amount of time to mark, comment on and turn-around student work. The increases in cognitive load which accompany this pressure may therefore lead a lecturer to adopt a more schema-driven approach (Bargh and Thein 1985; Plessner 2005). This contention is supported by Snyder and Stukas (1999) who found that, under conditions of high cognitive load, perceivers attempt to manage the task of interpreting information by placing an increased reliance on their expectancies; as opposed to attending to individuating information. However, this may also lead lecturers to bias their information processing in line with their expectancies. As a result, self-fulfilling prophecies, perceptual biases and cognitive dissonance might all have a considerable impact upon the marking process.

The participants in this study were not constrained by time, however, but were instead asked to mark the two essays at their earliest convenience. Consequently, they are likely to have engaged in a more data-driven information processing approach (Anderson 1981). Indeed, Fiske and Neuberg (1990) argued that, when motivated and in possession of sufficient attentional resources, individuals are more likely to apply a more data-driven processing strategy. Furthermore, Pendry and Macrae (1996) found that when perceivers were motivated to form an accurate perception of a target person, they were more likely to use a data-driven information processing approach. Given that self-fulfilling prophecies, perceptual biases and cognitive dissonance are all less likely when a data-driven approach is adopted, the time constraints placed on the participants might well explain the conflicting results of this study and those of previous research conducted in more naturalistic settings.

However, the expectancies imposed on the participants in the present study were also based on artificial information. Although expectancies can be derived from both indirect and direct personal experience, White, Jones, and Sherman (1998) argued that the extent to which information derived from indirect experience influences expectancy formation is determined by the degree of credibility the perceiver assigns to the source of such information. Bradley (1984) also suggested that repeated interactions with an individual may create a stronger expectation with regard to the quality of student work. This has the potential to either intensify or eliminate bias, depending on the accuracy of the judgement. Thus, future research may well need to consider the extent of personal knowledge a lecturer has about a student when marking their assessment, as well as the perceived likelihood of any future interactions between the lecturer and student.

With regard to the written feedback provided, Jussim (1986) argued that the teacher provides clearer feedback to the low-performing high-expectancy students as a means to bring them closer to their expectation, whereas the high-performing low-expectancy students receive less intricate support. However, in the present study, there does not seem to be a noticeable difference with regard to the clarity or intricacy of the feedback comments provided by the participants across the three reputation conditions. Instead, the most notable difference between the reputation conditions would appear to be in the increased amount of feedback provided on the negative student essay. This may be a reflection of higher education today; whereby academic staff are possibly too concerned with 'bringing the tail up', at the potential detriment

of their stronger students (Sadler 2010).

However, there would also seem to be a general imbalance between the number of feedback comments pertaining to the relative strengths and suggested areas of improvement across both the neutral and negative reputation conditions. Indeed, whilst the neutral reputation condition contained more comments concerning areas of improvement than strengths (13 and 2, respectively), the negative reputation condition elicited more comments about the strengths of the assignment than suggested areas of improvement (16 and 12, respectively). Conversely, there was a general balance between the number of comments pertaining to the relative strengths and suggested areas of improvement (5 and 6, respectively) within the positive reputation condition. Given that feedback can have both positive and negative behavioural consequences in terms of the effort, persistence, attention, participation and cooperation students are willing to put into future assessments (Jussim 1986, 1989; Jussim and Eccles 1992), additional research is clearly needed to examine the potential influence of expectancy-induced feedback bias on student behaviour. Yet, the biases in feedback construction observed in the present study do tentatively suggest that negative expectancy effects may be more powerful than positive expectancy effects within the marking process.

The observations of Sutherland and Goldschmid (1974) lend partial support to this contention in that these researchers also identified negative expectancy effects to generally be more powerful than positive ones. However, the findings of the present study should be interpreted with caution since negative expectancy effects only exerted a more powerful influence over the markers with regard to the amount of feedback provided. Indeed, there were no apparent differences in the marks awarded to student work, or the clarity or intricacy of the feedback comments provided across the three reputation conditions. Furthermore, Babad, Inbar, and Rosenthal (1982) found there to be no consistent trends regarding the strength of positive and negative expectancies, and Madon, Jussim, and Eccles (1997) reported that positive expectancy effects were generally more powerful than negative ones. As a result, future research should consider the conditions under which the occurrence of both positive and negative expectancy effects are facilitated (Jussim and Harber 2005).

Moreover, although the total number of comments that were made on the negative student essay would appear to be higher than on the other two student reputation profiles, no comparison was made between the same markers on the control student essay. It is therefore difficult to ascertain whether the increased amount of feedback provided on the negative student essay is an artefact of marker behaviour or the manipulation in reputation information. Indeed, it could just be that the group of markers who were randomly divided into the negative student reputation condition generally make more comments on student work. An examination of the impact of reputation information on feedback construction within, and between markers, on a number of different essays, would therefore be an important methodological consideration within future research.

From a psycholinguistics stance, Huot (1990) has also argued that reading is a fluid process, whereby the reader's response is often the result of their expectations. How a text is received and accepted is therefore somewhat predetermined, implying that the lecturer may see qualities in the essay which complement their expectation, but are not in fact present or valid. Thus, the teacher's expectation shapes their experience of reading a student's essay, which leads to the provision of marks that do not accurately represent the true merit of the work (Brennan 2008). However, whilst this would appear to be a plausible explanation for the results of previous research, the fact that all of the participants in the present study were aware that their feedback would be looked at by their peers may have encouraged them to invest more time and effort in their responses. This is an Accepted Manuscript of an article published by Taylor & Francis in ASSESSMENT & EVALUATION IN HIGHER EDUCATION on 9 December 2011, available online.

such, psycholinguistic biases may have had a limited effect on the results of this study.

In addition, only one marker explicitly acknowledged the reputation information provided. They commented that 'You have made a real effort to improve your writing style ...' (comment on negative profile – marked at 64%). The use of the word improve would indicate that the marker was comparing this essay to a previous attempt. However, as no such information was available, it can only be assumed that the negative student reputation profile was the comparison. Although the procedure utilised within the present study ensured that the reputation information was acknowledged by all of the participants (Jones, Paull, and Erskine 2002), this was the only comment which addressed the previous achievements of the student, or in this case, lack of. It might therefore be argued that the participants in the present study attempted to approach each task with a 'clean slate'.

However, implicit expectancies (i.e. those expectancies which are formed outside of the consciousness of the perceiver) can still impact an individual's responses – even when that individual is unaware of such expectancies (Chen and Bargh 1997; Bargh 2006; McCulloch et al. 2008). Such evidence has important implications for the extent to which the consequences of interpersonal expectancies can be harnessed and/or prevented. Indeed, if expectancies are explicit (i.e. formed consciously), they can be more easily identified and encouraged (or challenged) than those expectancies that are implicit and thus, more difficult to recognise (Wiers et al. 2005). However, by increasing a lecturer's awareness of the expectancies they hold, and their potential to impact the marking process, a lecturer might still be able to avoid those biases which emanate from self-fulfilling prophecies, cognitive dissonance and expectancy effects in general.

Pedagogical implications

The National Union of Students (NUS) has campaigned for anonymous marking since 1999, arguing that it provides universities with one remedial method against perceived discrimination. Brennan (2008) further argued that anonymous marking can help to reassure students that any concerns regarding one assessment can be voiced without fear of a tutor's backlash on future submissions. In addition, it is believed that such a system shifts the responsibility for learning towards the student, whereby they are expected to follow up specific feedback and support (Whitelegg 2002). Ultimately, anonymous marking 'safeguards' both the staff and the student, with some going as far to state that it reduces the tension between the two parties, facilitating their relationship, which promotes learning (Brennan 2008).

Alternatively, it is possible that anonymous marking compromises the openness of the teacher–student relationship with neither side directly communicating, and thus it can encourage a climate of distrust. Student learning may also be impinged by a lack of personalised commentary provided throughout the assessment, which is highly valued by many students (Jessop 2007). This particular disadvantage is outlined by Whitelegg (2002), who regards such marking procedures as a 'disruption to the feedback loop' (7) and promoting a homogenous view of the student body. As a consequence, weaker students can easily go undetected and are less likely to receive the essential support they require, and so the system can in fact discriminate against those it was designed to protect (Whitelegg 2002).

Issues of practicality are also raised, not just in terms of the increased administrative workload and error it entails, or the increased amount of time it will take to turnaround student work, but also the difficulty higher education institutions would have with implementing anonymous marking across the board. Indeed, not all

ment formats, such as those which require the tutor to directly observe the student (Owen, Stefaniak, and Corrigna 2010). It is also apparent that any decisions regarding student anonymity in assessment cannot be made without either side of the academic divide being disadvantaged. Nonetheless, the quantitative results of the present study should help to re-assure the student body about the quality of the systems by which their knowledge is developed and judged.

However, future research examining the influence of expectancy-induced biases on the marks awarded to student work will need to be undertaken before any such claims can be substantiated. In particular, future research will need to consider the extent to which the study fully replicates naturalistic circumstances. For instance, the time–pressures associated with marking will need to be accounted for, as will the perceived likelihood of any future interactions between the lecturer and student, and the extent of direct personal knowledge a lecturer has about a student when marking their assessment. Moreover, although the mixed-methodological approach adopted within the present study restricted the authors from doing so, future research should also look to examine potential marker bias within a much larger sample. Indeed, in order to achieve an adequate level of power (Cohen 1988, 1992) and a medium effect size, future research using a similar experimental design to the one described herein would require approximately 150 participants (G*Power [online]).

Additional research is also needed before the existence of expectancy-induced biases in summative feedback can be either confirmed/rejected. For example, a qualitative analysis of the feedback comments provided by a single marker, on a number of essays of a similar standard, could be undertaken. This would give future researchers a direct comparison of feedback, relating only to the profile of the student, as opposed to those differences in feedback caused by the personal preferences and idiosyncrasies of the marker; which may blur the answers being explored. The examination of how such feedback is constructed throughout an essay, and the behavioural and affective consequences of such feedback, should provide a fruitful avenue for future research. However, future research examining the extent to which audio and/or video feedback might be affected by lecturer expectancies would also be an interesting avenue for future research. Nevertheless, it is hoped that the results of the present study will, at the very least, stimulate further discussion about alternative means to reduce the perception of bias in marking.

Conclusion

Knowledge of the processes that underpin bias in marking is necessary if higher education institutions are to generate the means to counteract and prevent discrimination in marking. Unfortunately, as a consequence of the equivocal results thus far, many investigations maintain a focus on determining if biases are in operation, as opposed to accounting for their presence. Yet, there is value in reviewing how they may originate within the marking process. Much of the theory which can be applied to bias in marking is intertwined, and collectively offers researchers with a framework from which to examine potential marker bias in experiments. However, in order to confidently apply these theories to the context of marking in higher education, further empirical testing is required to confirm their operation. Thus, many of the propositions made above remain as hypothetical possibilities.

In addition, although the results of the present study do have a number of potential implications for the ongoing anonymous marking debate, the primary aim of this study was to explore the influence of reputation information on the assessment of undergraduate student work, and not to solely address the complex issues associated with anonymous marking. As a result, the reliance placed upon such findings

when there seems to be a lot of pressure from across the sector to move towards a uniform model whereby all student work is anonymously marked, the results of the present study, which failed to identify any real need for anonymous marking, could be used as a form of evidence (albeit limited) to help higher education institutions to resist this necessity. However, future research not only needs to consider those expectancy-induced biases which may be present in the marks awarded to student work, but also to examine the impact of such biases on the written feedback provided to students.

Notes on contributors

John Batten is a lecturer in the Sports Studies Department at the University of Winchester. His present research interests generally operate within the quantitative paradigm and have examined expectancy effects in both marking and sport.

Jo Batey is a senior lecturer in the Sports Studies Department at the University of Winchester. She has published in the areas of pedagogical cultural change, the exercise experiences of female runners, and the impact of career-ending injury and loss of the muscled self in bodybuilding.

Laura Shafe is a research assistant in the Sports Studies Department at the University of Winchester. Her research interests include relapse from physical activity and the psychology of injury.

Laura Gubby is a PhD student at Canterbury Christ Church University. Her doctoral thesis involves a sociological investigation into korfbal.

Phil Birch is a PhD student in Sport and Exercise Psychology at the University of Chichester. His doctoral thesis is examining measurement issues in mental toughness.

References

- Anderson, N.H. 1981. Foundations of information integration theory. San Diego, CA: Academic Press.
- Archer, J., and B. McCarthy. 1988. Personal biases in student assessment. *Educational Research* 30, no. 2: 142–4.
- Babad, E.Y., J. Inbar, and R. Rosenthal. 1982. Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers. *Journal of Educational Psychology* 74, no. 4: 459–74.
- Baird, J. 1998. What's in a name? Experiments with blind marking in A-level examinations. *Educational Research* 40, no. 2: 191–202.
- Bargh, J.A. 2006. What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behaviour. *European Journal of Social Psychology* 36, no. 2: 147–68.
- Bargh, J.A., and R.D. Thein. 1985. Individual construct accessibility, person memory, and the recall-judgment link: The case of information overload. *Journal of Personality and Social Psychology* 49: 1129–46.
- Bradley, C. 1984. Sex bias in the evaluation of students. *British Journal of Social Psychology* 23, no. 2: 147–53.
- Bradley, C. 1993. Sex bias in assessment overlooked. *Assessment & Evaluation in Higher Education* 18, no. 1: 3–8.
- Braun, C. 1976. Teacher expectation: Social-psychological dynamics. *Review of Educational Research* 46, no. 2: 185–213.
- Brennan, D.J. 2008. University student anonymity in the summative assessment of written work. *Higher Education Research & Development* 27, no. 1: 43–54.
- Brophey, J.E. 1983. Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology* 75, no. 5: 631–61.
- Chapman, L.J., and J.P. Chapman. 1967. Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology* 72: 193–204.
- Chen, M., and J.A. Bargh. 1997. Nonconscious behavioural confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology* 33: 541–60.
- Cohen, J. 1988. Statistical power for the behavioural sciences. New York, NY: Academic Press.
- Cohen, J. 1992. A power primer. *Psychological Bulletin* 112, no. 1: 155–9.
- Dennis, I., and D.E. Newstead. 1994. The strange case of the disappearing sex bias. *Assessment & Evaluation in Higher Education* 19, no. 1: 49–56.
- Dennis, I., S.E. Newstead, and D.E. Wright. 1996. A new approach to exploring biases in educational assessment. *British Journal of Psychology* 87: 515–34.
- Diederich, P.B. 1974. Measuring growth in English. Urbana, IL: National Council of Teachers of English.
- Dipboye, R.L. 1985. Some neglected variables in research on discrimination in appraisals. *Academy of Management Review* 10: 116–27.
- Ecclestone, K. 2001. 'I know a 2:1 when I see it': Understanding criteria for degree classifications in Franchised University programmes. *Journal of Further and Higher Education* 25, no. 3: 301–13.
- Elander, J. 2002. Developing aspect-specific assessment criteria for essays and examination answers in psychology. *Psychology Training Review* 10: 31–51.
- Elander, J., K. Harrington, L. Norton, H. Robinson, and P. Reddy. 2006. Complex skills and academic writing: A review of evidence about the types of learning required to meet core assessment criteria. *Assessment & Evaluation in Higher Education* 31, no. 1: 71–90.
- Festinger, L. 1957. A theory of cognitive dissonance. Stanford, CA: Stanford University Press.
- Field, A. 2005. Discovering statistics using SPSS. 2nd ed. London: Sage.
- Fiske, S.T., and S.L. Neuberg. 1990. A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In *Advances in experimental social psychology*, ed. M.P. Zanna, 1–74. New York, NY: Academic Press.
- Fiske, S.T., and S.E. Taylor. 1991. Social cognition. 2nd ed. New York, NY: McGraw-Hill.
- Fleming, N. 1999. Biases in marking students' written work. In *Assessment matters in higher education*, ed. S. Brown and A. Glasner, 83–92. Buckingham: Open University Press.

- Francis, B., J. Robson, and B. Read. 2001. An analysis of undergraduate writing styles in the context of gender and achievement. *Studies in Higher Education* 26, no. 3: 313–26.
- Francis, B., J. Robson, and B. Read. 2002. Gendered patterns in writing and degree award. In *Gender, teaching and research in higher education: Challenges for the twenty-first century*, ed. G. Howie and A. Tauchert, 167–80. Hampshire: Ashgate Press.
- Francis, B., J. Robson, B. Read, and L. Melling. 2003. Lecturers' perceptions of gender and undergraduate writing style. *British Journal of Sociology of Education* 24, no. 3: 357–73.
- Gratton, C., and I. Jones. 2004. *Research methods for sport studies*. London: Routledge.
- Greenlees, I., H. Webb, B. Hall, and A. Manley. 2007. Curmudgeon or golden-ager?: Reported exercise participation influences the perception of older adults. *Journal of Sport & Exercise Psychology* 29: 333–47.
- G*Power [online]. General power analysis program. <http://www.psych.uni-duesseldorf.de/aap/projects/gpower/> (accessed 20/07/2011).
- Hand, L., and D. Clews. 2000. Marking the difference: An investigation of the criteria used for assessing undergraduate dissertations in a business school. *Assessment & Evaluation in Higher Education* 25 no. 1: 5–21.
- Higgins, E.T., and J.A. Bargh. 1987. Social cognition and social perception. *Annual Review of Psychology* 38: 369–425.
- Hilton, J.L., and J.M. Darley. 1991. The effects of interaction goals on person perception. In *Advances in experimental social psychology*, ed. M.P. Zanna, 235–67. London: Academic Press.
- Huot, B. 1990. The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research* 60, no. 2: 237–63.
- Jessop, T. 2007. Student perceptions of assessment and feedback. Unpublished report to the University Learning and Teaching Committee. The University of Winchester.
- Jones, M.V., G.C. Paull, and J. Erskine. 2002. The impact of a team's aggressive reputation on the decisions of Association Football referees. *Journal of Sports Sciences* 20: 991–1000.
- Jussim, L. 1986. Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review* 93, no. 4: 429–45.
- Jussim, L. 1989. Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology* 57, no. 3: 469–80.
- Jussim, L., and J.S. Eccles. 1992. Teacher expectations II: Construction and reflection of student achievement. *Journal of Personality and Social Psychology* 63, no. 6: 947–61.
- Jussim, L., and K.D. Harber. 2005. Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review* 9, no. 2: 131–55.
- Madon, S.J., L. Jussim, and J. Eccles. 1997. In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology* 72: 791–809.
- Martin Ginis, K.A., A.E. Latimer, and M.E. Jung. 2003. No pain no gain? Examining the generalizability of the exerciser stereotype to moderately active and excessively active targets. *Social Behaviour and Personality* 31: 283–90.
- Martin, K.A., A.R. Sinden, and J.C. Fleming. 2000. Inactivity may be hazardous to your image: The effects of exercise participation on impression formation. *Journal of Sport & Exercise Psychology* 22: 283–91.
- Maykut, P.S., and R.E. Morehouse. 1994. *Beginning qualitative research: A philosophical and practical guide*. London: Falmer Press.
- McCulloch, K.C., M.J. Ferguson, C.C.K. Kawada, and J.A. Bargh. 2008. Taking a closer look: On the operation of nonconscious impression formation. *Journal of Experimental Social Psychology* 44, no. 3: 614–23.
- Meadows, M., and L. Billington. 2005. A review of the literature on marking reliability. Paper commissioned by the National Assessment Agency.

- Merton, R.K. 1948. The self-fulfilling prophecy. *The Antioch Review* 8, no. 2: 193–210.
- Newstead, S.E. 1996. The psychology of student assessment. *The Psychologist* 9, no. 13: 543–7.
- Newstead, S., and I. Dennis. 1993. Bias in student assessment. *The Psychologist* 6, no. 10: 451–2.
- Newstead, S.E., and I. Dennis. 1990. Blind marking and sex bias in student assessment. *Assessment & Evaluation in Higher Education* 15, no. 2: 132–9.
- Olson, J.M., N.J. Roese, and M.P. Zanna. 1996. Expectancies. In *Social psychology: Handbook of basic principles*, ed. E.T. Higgins and A.W. Kruglanski, 211–38. New York, NY: The Guildford Press.
- Owen, C., J. Stefaniak, and G. Corrigan. 2010. Marking identifiable scripts: Following up student concerns. *Assessment & Evaluation in Higher Education* 35, no. 1: 33–40.
- Pain, R., and G. Mowl. 1996. Improving geography essay writing using innovative assessment. *Journal of Geography in Higher Education* 20, no. 1: 19–31.
- Pendry, L.F., and C.N. Macrae. 1996. What the disinterested perceiver overlooks: Goal directed social categorization. *Personality & Social Psychology Bulletin* 22: 249–56.
- Plessner, H. 2005. Positive and negative effects of prior knowledge on referee decisions in sport. In *The routines of decision making*, ed. T. Betsch and S. Haberstroh, 311–24. Mahwah, NJ: Lawrence Erlbaum.
- Read, B., B. Francis, and J. Robson. 2005. Gender, ‘bias’, assessment and feedback: Analyzing the written assessment of undergraduate History essays. *Assessment & Evaluation in Higher Education* 30, no. 3: 241–60.
- Rigsby, L.C. 1987. Changes in students’ writing and their significance. Paper presented at the meeting of the Conference on College Composition and Communication, Atlanta.
- Rust, C. 2002. The impact of assessment on student learning: How can the research literature practically help inform the development of departmental assessment strategies and learner-centred assessment practices? *Active Learning in Higher Education* 3, no. 2: 145–58.
- Sadler, D.R. 2010. Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education* 35, no. 5: 535–50.
- Snyder, M., and A.A. Stukas, Jr. 1999. Interpersonal processes: The interplay of cognitive, motivational, and behavioural activities in social interaction. *Annual Review of Psychology* 50: 273–303.
- Spear, M. 1984. Sex bias in science teachers’ ratings of work and pupil characteristics. *European Journal of Science Education* 6, no. 4: 369–77.
- Sutherland, A., and M.L. Goldschmid. 1974. Negative teacher expectation and IQ change in children with superior intellectual potential. *Child Development* 45: 852–6.
- Tabachnick, B.G., and L.S. Fidell. 2007. *Using multivariate statistics*. Boston, MA: Allyn and Bacon.
- White, K.J., K. Jones, and M.D. Sherman. 1998. Reputation information and teacher feedback: Their influences on children’s perceptions of behaviour problem peers. *Journal of Social & Clinical Psychology* 17: 11–37.
- Whitelegg, D. 2002. Breaking the feedback loop: Problems with anonymous assessment. *Planet* 3: 7–8.
- Wiers, R.W., J. van de Luitgaarden, E. van den Wildenberg, and F.T.Y. Smulders. 2005. Challenging implicit and explicit alcohol-related cognitions in young heavy drinkers. *Addiction* 100: 806–19.